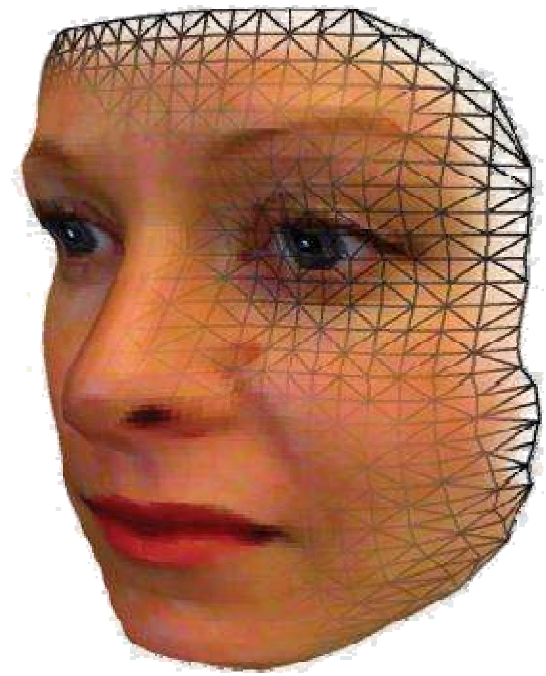
3D Face Recognition
Approaches and Challenges



Amirhosein Nabatchian

August 2008

Outline

- Research objective
- Introduction
- Literature review
- Challenges

Research Objective

- Face recognition:

  Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces

Background

- Face Authentication/Verification (1:1 matching)

- Face Identification/Recognition (1:N matching)

- Procedure

Image Acquisition

Capture face images & Generate 3D models

→

Pre-processing

Normalize images into the same position

→

Feature Extraction

Extract the features from normalized face images

→

Classification

Design a classifier, train it with dataset, and test its validity

Types of 3D Data

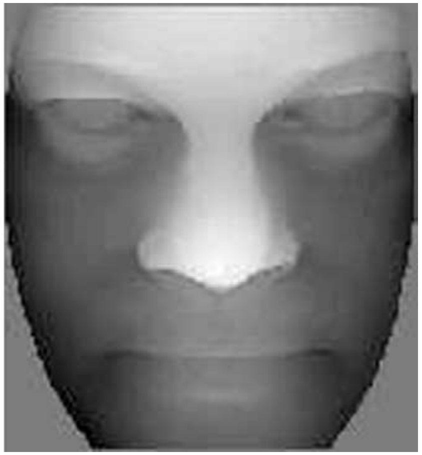
- *Point-Cloud Representation*
  - The point cloud is the set of the 3-D coordinates *{x, y, z}* of the points of a face object. A face with *N* samples is simply represented in terms of three coordinate vectors *X, Y , and Z of* length *N.*
- *Range Image*
  - the z-coordinates of the face points *are* mapped on a regular x–y grid by using linear interpolation.
  - has the form of a 2-D function *I(x, y),* similar to an intensity image, so it's simple .
  - Invariant to the change of illumination & color

- Surface-normal based
  - each point of the facial point-cloud data is described by its 3-D (nx, ny, nz) unit normal vector.
- *Curvature-Based Representation*
  - are invariant to rotations
  - three-vector and their derivatives, i.e., the mean (H) and Gaussian (K) curvatures extracted from each facial surface point.
  - mean curvature- and Gaussian curvature-based representations

- *3-D Voxel Representation*
  - point-cloud data is converted to a voxel structure, denoted as *Vd(x, y, z), by imposing a lattice.*

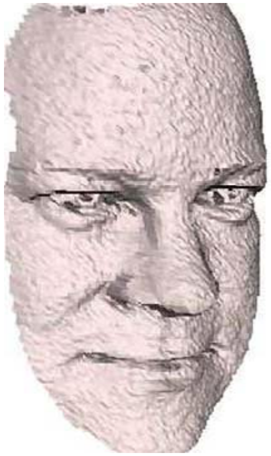
3D Samples

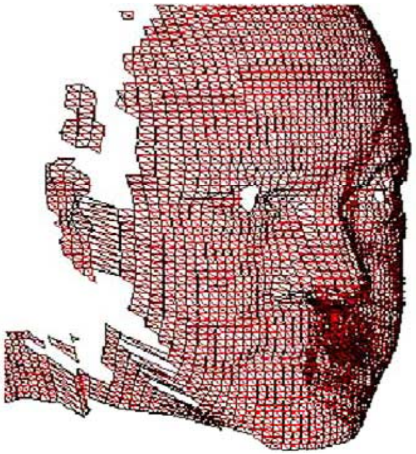
Cropped 2D intensity image



3D range image

3D shaded curvature model





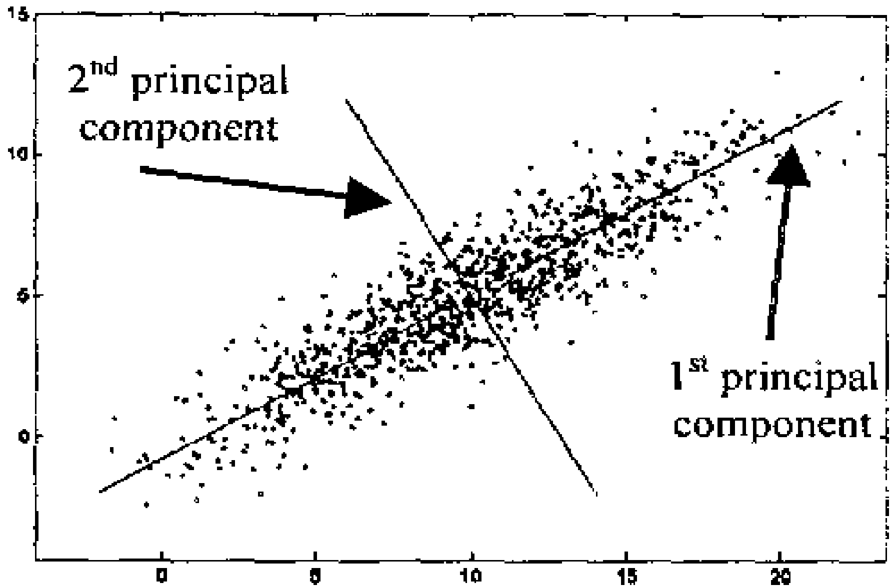
3D Mesh
3d Voxel

FRGC Dataset

FRGC: Face Recognition Grand challenge
-A NIST program
-Emotions: neutral, angry, happy, sad, surprised, disgusted, and puffy

PCA

- Principle Component Analysis (PCA) to reduce the dimensionality
- Principal components are eigenvectors of covariance matrix



PCA (cont)

- Modeling
  1. Given a collection of *n* labeled training images,
  2. Compute mean image and covariance matrix. Subtract the mean.
  3. Compute k Eigenvectors (note that these are images) of covariance matrix corresponding to k largest Eigenvalues.
  4. Project the training images to the k-dimensional Eigenspace.
- Recognition
  1. Given a test image, project to Eigenspace.
  2. Perform classification to the projected training images.

PCA vs LDA

- Between-class scatter

$$S_B = \sum_{i=1}^{c} |\chi_i| (\mu_i - \mu)(\mu_i - \mu)^T$$

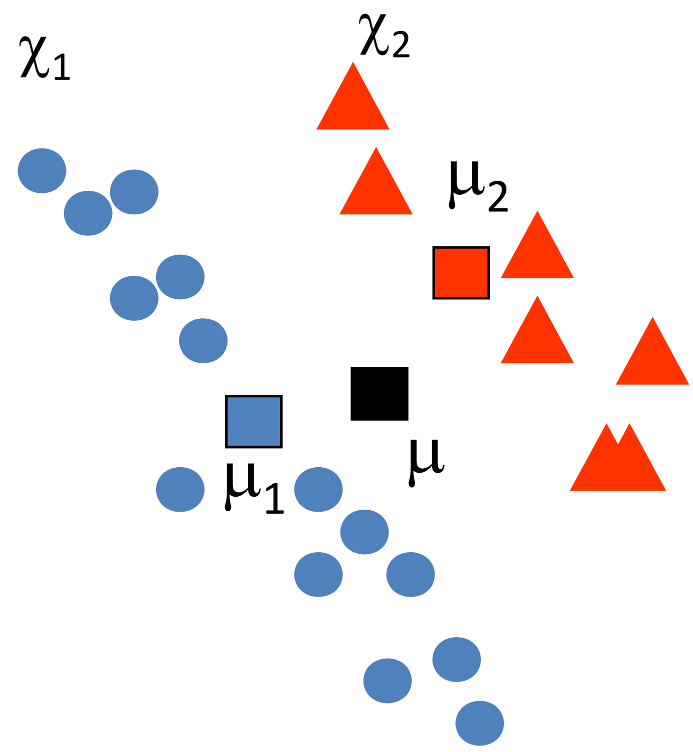
- Within-class scatter

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu_i)(\mu_k - \mu_i)^T$$

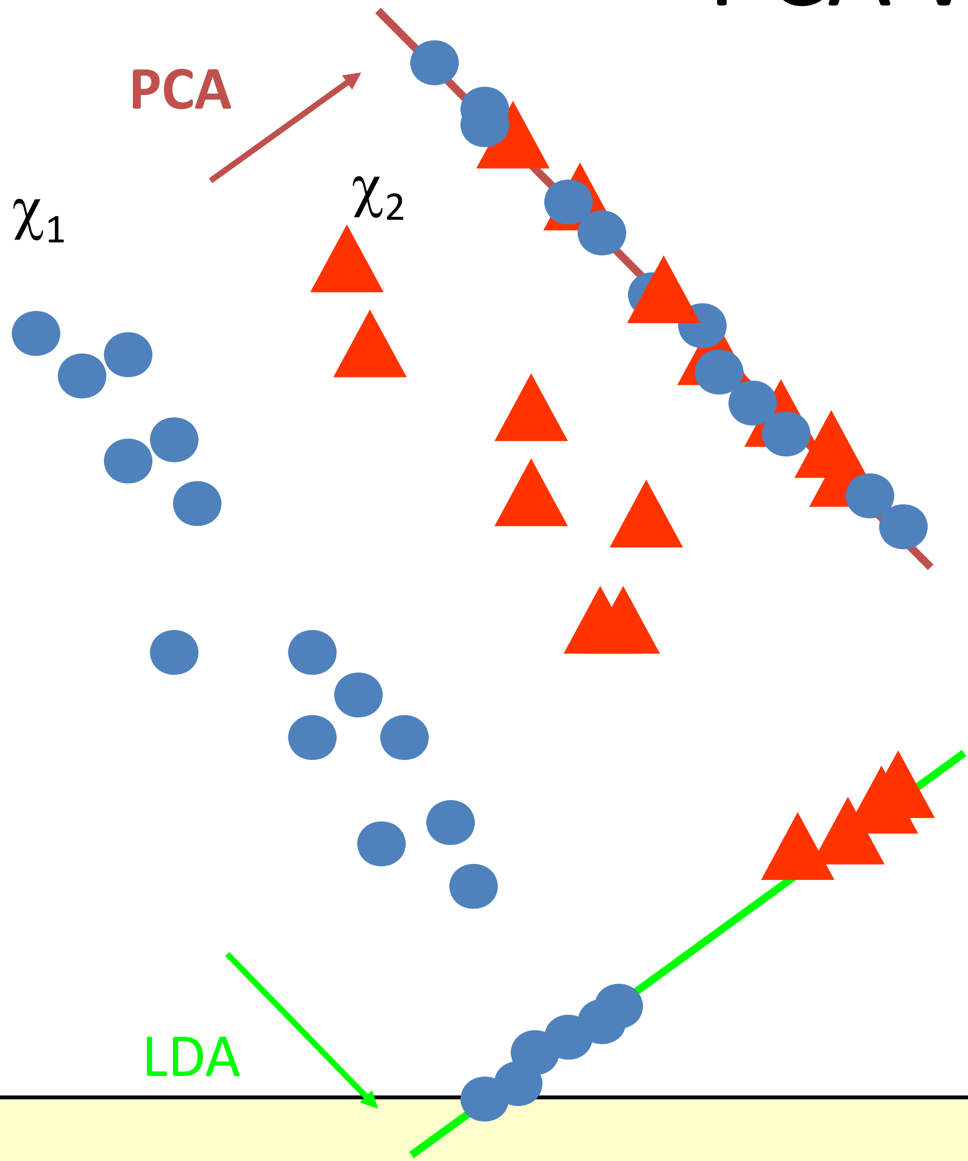
- Total scatter

$$S_T = \sum_{i=1}^{c} \sum_{x_k \in \chi_i} (x_k - \mu)(\mu_k - \mu)^T = S_B + S_W$$

- Where
  - c is the number of classes
  - μ_i is the mean of class χ_i
  - $|\chi_i|$ is number of samples of χ_i.



PCA vs LDA



- PCA (Eigenfaces)

$$W_{PCA} = \arg\max_{W} \left| W^T S_T W \right|$$

Maximizes projected total scatter

- Fisher's Linear Discriminant

$$W_{fld} = \arg\max_{W} \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|}$$

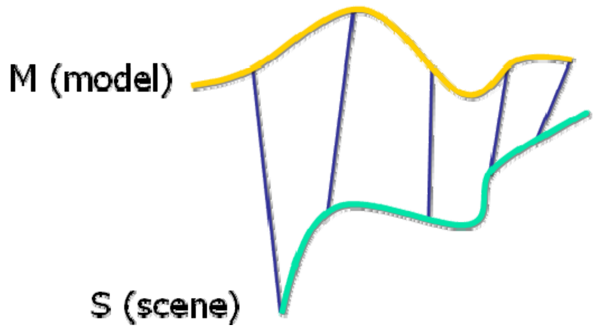
Maximizes ratio of projected between-class to projected within-class scatter

ICP

- **Iterative Closest Point** (**ICP**) is an algorithm employed to match two clouds of points. This matching is used to reconstruct 3D surfaces from different scans, to localize robots, etc.
- The algorithm is very simple and is commonly used in real-time. It iteratively estimates the transformation (translation, rotation) between two raw scans.
- Inputs: two raw scans, initial estimation of the transformation, criteria for stopping the iteration.
- Output: refined transformation.
- Essentially the algorithm steps are:
- Associate points by the nearest neighbor criteria.
- Estimate the parameters using a mean square cost function.
- Transform the points using the estimated parameters.
- Iterate (re-associate the points and so on).

ICP

- Let M be a model point set.
- Let S be a scene point set.



We assume :

1. $N_M = N_S$.
2. Each point S_i correspond to M_i .

ICP

The MSE objective function :

$$f(R,T) = \frac{1}{N_S} \sum_{i=1}^{N_S} \left\| m_i - Rot(s_i) - Trans \right\|^2$$

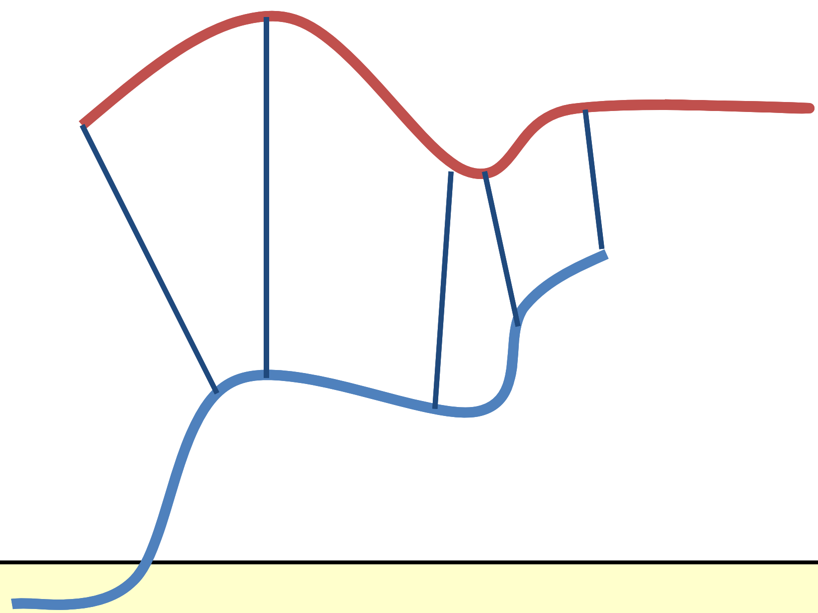
$$f(q) = \frac{1}{N_S} \sum_{i=1}^{N_S} \left\| m_i - R(q_R) s_i - q_T \right\|^2$$

The alignment is :

$$(rot, trans, d_{mse}) = \Phi(M, S)$$

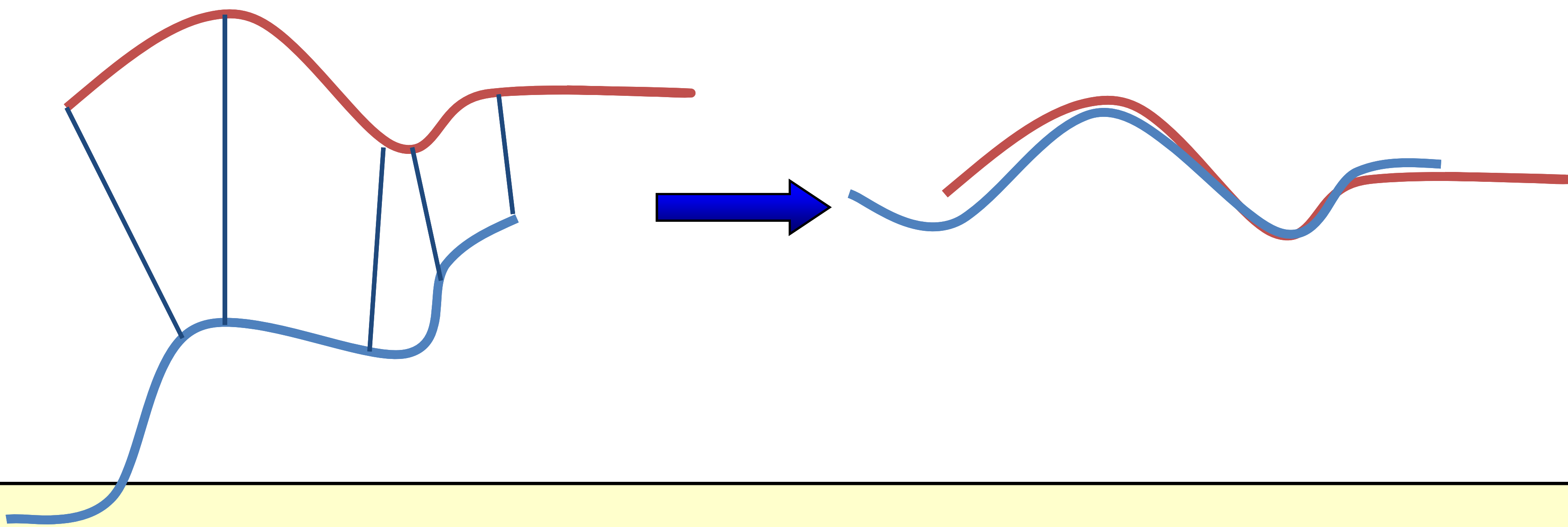
Aligning 3D Data

- How to find correspondences: User input? Feature detection? Signatures?
- Alternative: assume closest points correspond



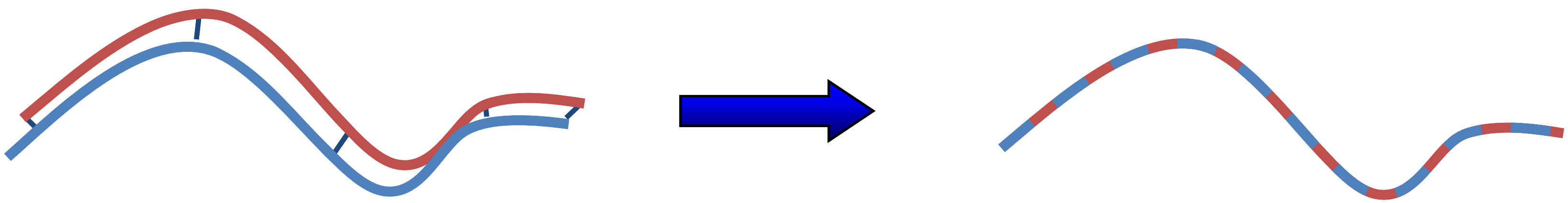
Aligning 3D Data

- How to find correspondences: User input? Feature detection? Signatures?
- Alternative: assume closest points correspond

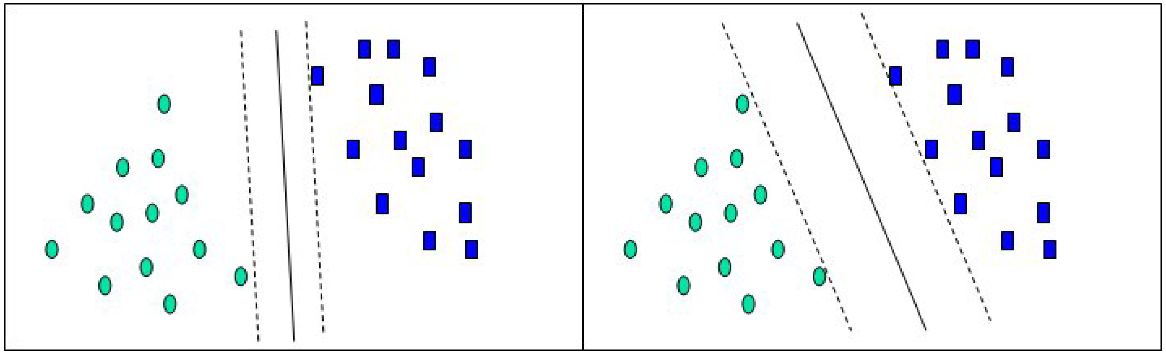


Aligning 3D Data

- Converges if starting position “close enough“



Support Vector Machines

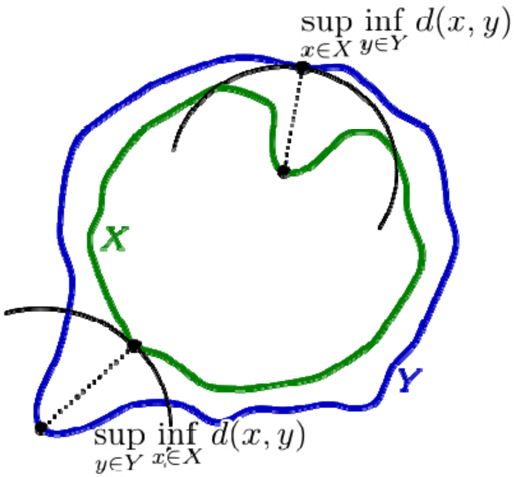


Small Margin Large Margin

- When not linearly separable, transform to a higher order space, separate linearly, and transform to the output space

Hausdorff distance

- measures how far two compact non-empty subsets of a metric space are from each other.
- Let *X* and *Y* be two compact subsets of a metric space *M*. The Hausdorff distance $d_H(X,Y)$ is the minimal number *r* such that the closed *r*-neighborhood of any *x* in *X* contains at least one point *y* of *Y* and vice versa. In other words, if *d(x, y)* denotes the distance in *M*, then



$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y),\ \sup_{y\in Y}\inf_{x\in X} d(x,y)\},$$

- Cartoux (1989)

Method	1. segmenting a range image based on principal curvature 2. Finding a plane of bilateral symmetry through the face
Remark	consider methods of matching the profile from the plane of symmetry
Advantage	100% Accuracy
Disadvantage	Small dataset of 6

- Lee (1990)

Method	1. segment convex regions in a range im-age based on the sign of the mean and Gaussian curvatures 2. create an extended Gaussian image (EGI) for each
Remark	EGI, Matching by correlation
Advantage	Less changes due to facial expression
Disadvantage	Not sensing object size

- Gordon (1991)

Method	1. Calculate principal curvatures on the surface 2. Generate face descriptors from curvarture map
Remark	Outline of the use of curvature information in the process of face recognition
Advantage	Can deal with faces different in size
Disadvantage	Need some extension to cope with changes in facial expression

- Spherical Correlation [Tanaka & Ikeda (1998)]

Method	1. Construct Extended Gaussian Image (EGI) 2. Compute Fisher's spherical correlation on EGI's
Remark	First work to investigate and evaluate free-formed curved surface recognition
Advantage	Simple, efficient, and robust to distractions such as glasses and facial hair
Disadvantage	Not tested on faces in different sizes and facial expressions

- Eigenface [Achermann et al. (1997)]

Method	1. Consider face images as vectors 2. Apply principal component analysis (PCA)
Remark	1. Optimal in the least mean square error sense 2. Prevalent method in 2D face recognition [Turk & Pentland (1991)]
Advantage	Large dimension reduction
Disadvantage	Bad performance with large database

- Optimal Linear Component [Liu et al. (2004)]

Method	1. Consider face images as vectors 2. Find optimal linear subspaces for recognition
Remark	Optimal in the sense that the ratio of the between-class distance and within-class distance is maximized
Advantage	Better performance than standard projections, such as PCA, ICA, or FDA
Disadvantage	Lots of computation due to optimization problem

- Fusion [Gokberk (2005)]

Method	Fusing Gaussian images, ICP matching, range profile, PCA, and LDA
Remark	explore methods of fusing the results of the five approaches
Advantage	Good results
Disadvantage	More Computation

- Lee (2005)

Method	1- Curvature values at 8 points 2- SVM
Remark	explore methods of fusing the results of the five approaches
Advantage	use a Cyberware sensor to acquire the enrollment images
Disadvantage	feature points are manually located.

- multi-region method [Chang (2007)]

Method	multiple overlapping subregions around the nose are independently matched using ICP and the results are fused
Remark	Over 4000 images from over 400 persons
Advantage	easy , Large dataset

Recognition Algorithms and Their Results

Author, Year	Persons in dataset	Images in dataset	Image size	3d face data	Matching algorithm	Reported performance
Cartoux, 1989	5	18	Not available	Profile, Surface	Minimum distance	100%
Lee, 1990	6	6	256x150	EGI	Correlation	None
Gordon, 1992	26	26	Not available	Feature vector	Closest vector	100%
Nagamine, 1992	16	160	256x240	Multiple profiles	Closest vector	100%
Achermann, 1997	24	240	75x150	Range image	PCA, HMM	100%
Tanaka, 1998	37	37	256x256	EGI	Correlation	100%
Achermann, 2000	24	240	75x150	Point set	Hausdorff distance	100%
Chua, 2000	6	24	Not available	Point set	Point signature	100%
Hasher, 2003	37	222	242x347	Range image	PCA	97%
Lee, 2003	35	70	320x320	Feature vector	Closest vector	94%
Medioni, 2003	100	700	Not available	Point set	ICP	98%
Moreno, 2003	60	420	2.2 K points	Feature vector	Closest vector	78%
Pan, 2003	30	360	3 K points	Point set, Range	Hausdorff and PCA	3-5% EER
Lee, 2004	42	84	240x320	Range image	Weighted Hausdorff	98%
Lu, 2004	18	113	240x320	Point set	ICP	96%
Russ, 2004	200 FRGC V1	468	480x640	Range image	Hausdorff distance	98%
Xu, 2004	120	720	Not available	Point set, Feature vector	Minimum distance	72%
Bronstein, 2005	30	220	Not available	Point set	Canonical forms	100%
Chang, 2005	466 FRGC V2	4007	480x640	Point set	Multi-ICP	92%
Gökerberk, 2005	106	579	Not available	Multiple	Multiple	99%
Lee, 2005	100	200	Various	Feature vector	SVM	96%
Lu, 2005	100	196	240x320	Surface Mesh	ICP, TPS	89%
Pan, 2005	276 FRGC V1	943	480x640	Range image	PCA	95%
Passalis, 2005	466 FRGC V2	4007	480x640	Surface Mesh	Deformable model	90%
Russ, 2005	200 FRGC V1	398	480x640	Range image	Hausdorff distance	89.5%

Challenges

- Sensors
- illumination invariance
- Active vs passive

Sensors

- Passive stereo
  - two cameras with a known geometric relationship are used
- Pure structured light
  - uses a camera and a light projector with a known geometric relationship. A light pattern is projected into the scene, detected in an image acquired by the camera
- hybrid of passive stereo and structured lighting
  - a pattern is projected onto the scene and then imaged by a stereo camera rig

sensors

- Even under ideal illumination, it is common for artifacts to occur in face regions such as oily regions that appear specular, the eyes, and regions of facial hair such as eyebrows, mustache, or beard
- ‘‘holes’’ missing data
- ‘‘spikes’’ outlier error in the data, for example from an inter-reflection
- Depth of field for sensing (.3m for stereo , 1m for structured)
- Image acquisition time

- A 3D shape is illumination invariant
- Making the 3D image from 2D sensors is not

- If we have active sensing then acquiring time increases and movements of the object make it noisy

References

- [1] Kevin W. Bowyer *, Kyong Chang, Patrick Flynn , A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition , computer vision and image understanding (2006) 1-15
- [2] B. Achermann, H. Bunke, Classifying range images of human faces with Hausdorff distance, in: 15-th International Conference on Pattern Recognition, September 2000, pp. 809–813.
- [3] B. Achermann, X. Jiang, H. Bunke, Face recognition using range images, International Conference on Virtual Systems and MultiMedia (1997) 129–136.
- [4] C. Beumier, M. Acheroy, Face verification from 3D and grey level cues, Pattern Recognition Letters 22 (2001) 1321–1329.
- [5] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intel-ligence 25 (2003) 1063–1074.
- [6] C. Boehnen, P.J. Flynn, Accuracy of 3D scanning technologies in a face scanning context, in: Fifth International Conference on 3D Imaging and Modeling (3DIM 2005), June 2005, pp. 310–317.
- [7] K.W. Bowyer, Face recognition technology and the security versus privacy tradeoff, IEEE Technology and Society (2004) 9–20.
- [8] K.W. Bowyer, K. Chang, P.J. Flynn, A survey of 3D and multi-modal 3D + 2D face recognition, in: 17-th International Conference on Pattern Recognition, August 2004, pp. 358–
- 361.
- [9] K.W. Bowyer, K. Chang, P.J. Flynn, A survey of 3D and multi-modal 3D + 2D face recognition, Face Processing: Advanced Mod-eling and Methods, to appear.
- [10] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Expression-invariant 3D face recognition, in: International Conference on Audio-and Video-Based Person Authentication (AVBPA 2003), LNCS, vol. 2688, 2003, pp. 62–70.
- [11] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Three-dimensional face recognition, International Journal of Computer Vision (2005) 5–
- 30.
- [12] J.Y. Cartoux, J.T. LaPreste, M. Richetin, Face authentication or recognition by profile extraction from range images, in: Proceedings of the Workshop on Interpretation of 3D Scenes, 1989, pp. 194–
- 199.

- [13] K. Chang, K. Bowyer, P. Flynn, An evaluation of multi-modal 2D + 3D face biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4) (2005) 619–624.
- [14] K. Chang, K. Bowyer, P. Flynn, Face recognition using 2D and 3D facial data, in: Multimodal User Authentication Workshop, Decem-ber 2003, pp. 25–32.
- [15] K. Chang, K.W. Bowyer, S. Sarkar, B. Victor, Comparison and combination of ear and face images for appearance-based biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 25
- (9) (2003) 1160–1165.
- [16] K.I. Chang, K.W. Bowyer, P.J. Flynn, Adaptive rigid multi-region selection for handling expression variation in 3D face recognition, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.
- [17] C. Chua, F. Han, Y.K. Ho, 3D human face recognition using point signature, IEEE International Conference on Automatic Face and Gesture Recognition (2000) 233–238.
- [18] L. Farkas, Anthropometry of the Head and Face, Raven Press, New York, 1994.
- [19] A. Godil, S. Ressler, P. Grother, Face recognition using 3D facial shape and color map information: comparison and combination, in:
- Biometric Technology for Human Identification, SPIE, vol. 5404, April 2005, pp. 351–361.
- [20] B. Gokberk, A.A. Salah, L. Akarun, Rank-based decision fusion for 3D shape-based face recognition, in: International Conference on Audio-and Video-based Biometric Person Authentication (AVBPA 2005), LNCS, vol. 3546, July 2005, pp. 1019–1028.
- [21] G. Gordon, Face recognition based on depth and curvature features, Computer Vision and Pattern Recognition (CVPR) (June) (1992) 108–110.
- [22] C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range imaging, in: Seventh International Sympo-sium on Signal Processing and Its Applications, 2003, pp. 201–204.
- [23] M. Husken, M. Brauckmann, S. Gehlen, C. von der Malsburg, Strategies and benefits of fusion of 2D and 3D face recognition, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.
- [24] M.L. Koudelka, M.W. Koch, T.D. Russ, A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.